Remarks

Claims 1–3, 5–6, and 8–28 are pending in this application. Claims 1–3, 5–6, 9, and 11–15 have been amended to make editorial changes. Claims 4 and 7 have been canceled. New claims 22–28 have been added to more specifically claim the invention. No new matter has been added. The new and amended claims are fully supported by the specification.

Double Patenting Rejection

Claims 1–3, 5–6, and 9–21 have been rejected for obviousness-type double patenting as being unpatentable over claims 1–26 of U.S. patent 6,865,536 in view of U.S. patent 5,054,085 (Meisel). Applicant may overcome this rejection by submitting a terminal disclaimer. However, applicant will defer submitting a terminal disclaimer until the claims are allowable apart from the obviousness-type double patenting rejection.

Section 103 Rejections

Claims 1–3 were rejected under section 103 as being unpatentable over U.S. patent 5,960,399 (Barclay) in view of Meisel. Claim 5 and 6 are rejected under section 103 as being unpatentable over Barclay in view of Meisel, and further in view of U.S. patent 6,216,104 (Moshfeghi). Claims 8–13 and 18–21 have been rejected under section 103 as being unpatentable over Barclay in view of Meisel, and further in view of U.S. patent 5,751,951 (Osborne). Claims 14–17 have been rejected under section 103 as being unpatentable over Barclay in view of Meisel, and further in view of Osborne and Moshfeghi. Reconsideration of the rejections and allowance of the claims are respectfully requested for the following reasons.

No Suggestion to Combine Barclay and Meisel

There is *no suggestion or motivation to combine* Meisel with Barclay. These references are very dissimilar and moreover, inconsistent with each another.

Barclay discusses a method to identify or verify who a person is based on the unique features in that person's speech. Barclay, abstract. Each person's speech is unique because each person has unique physical characteristics (e.g., throat, mouth, lips, teeth, and nasal cavity). Barclay extracts unique features of a person's speech, which Barclay refers to as cepstra, in order to identify or verify the person. Barclay, abstract.

In sharp contrast to Barclay, Meisel discusses a speech preprocessing method that reduces the uniqueness of speech between different speakers. In its abstract, Meisel states:

> Thus after the pre-processing performed by this invention, the *parameters would look much the same* for the same word *independent of speaker*. In this manner, *variations in the speech signal* caused by the physical makeup of a speaker's throat, mouth, lips, teeth, and nasal cavity would be, at least in part, *reduced by the pre-processing*.

Thus, after Meisel preprocesses some speech, there would be fewer, if any, unique characteristics of that speech upon which Barclay's method may use. When Meisel is combined with Barclay, Barclay would likely not be able to identify or verify a person.

Therefore, one having ordinary skill in the art would not combine Meisel with Barclay. Barclay's method depends on maintaining the unique characteristics of each person's speech, while Meisel removes differences in speech that distinguish different speakers. There is no motivation to combine these references, and the examiner has not made a *prima facie* case of obviousness. For at least this reason, claims 1–3, 5–6, and 8–21 and their dependents should be allowable.

No Reasonable Expectation for Success

There is *no reasonable expectation of success* that combining Meisel with Barclay would result in a system that processes speech in such a way to preserve the uniqueness and characteristics of a person's speech, as in the invention.

Barclay extracts cepstral features from speech and sends only these features (not the speech itself) to a server for processing. Meisel reduces the uniqueness of speech between different speakers. Each reference does not show or suggest processing speech in such a way to preserve the uniqueness and characteristics of a person's speech, and thus, the combination would not be show or suggest this either.

Therefore, one having ordinary skill in the art *would have no reasonable expectation for success* that by combining these two references one would obtain the invention. The examiner has not made a *prima facie* case of obviousness. For at least this additional reason, claims 1–3, 5–6, and 8–21 and their dependents should be allowable.

Combination Falls Short

Even if the references were combined, and there is no suggestion to do this for the reasons discussed above, the combination still falls short of the recited invention. The combination of the references does not show or suggest each and every limitation of the recited invention.

As will be discussed in more detail below, Barclay discusses processing speech on the client side to extract a set of cepstral features for the server. The server then processes only these cepstral features. Meisel discusses generating parameters for preprocessing.

The combination of Barclay and Meisel would be a method that applies an extra preprocessing step of Meisel to the extracted cepstral features. Since the preprocessing in Meisel reduces rather than preserves or enhances the uniqueness of speech, the combination of Meisel and Barclay would, at best, process only the extracted cepstral features.

The combination does not process encoded speech and does not show or suggest each and every limitation recited in the invention. Therefore, the claims should be allowable.

Claim 1

In particular, for example, claim 1 recites a client to (a) "*store* the audio speech in one or more buffers in a *raw uncompressed audio format*, each buffer comprising a portion of the received audio speech," and (b) "*encode* a buffer of the received audio speech before all of the audio speech is received."

No Storing the Speech Itself in Raw Format

The combination of Barclay and Meisel does not show a client having a capability to "*store the audio speech* in one or more buffers *in a raw uncompressed audio format*, each buffer comprising a portion of the received audio speech."

Barclay does not capture and does not store speech in raw uncompressed audio format. At column 5, lines 48–50, Barclay describes storing only the extracted, quantized cepstral features. At column 2, lines 3–4 and column 5, line 4, Barclay describes a client processing speech to extract cepstral features. At column 2, lines 16–25, Barclay describes time masking, frequency masking, volume, frequency range, and amplitude dynamic range as representing cepstral features. However, Barclay never describes representing the *content* of the speech. The

cepstral features are some processed formations from applying some mathematical operations on the speech. These cepstral features are not speech. Column 2, lines 4–7.

Quantizing the cepstral features does not convert these features back to speech. Barclay's client further processes these cepstral features by quantizing them. At column 5, lines 47–48, Barclay describes its client as quantizing "features *from* the raw digitized information data." Regardless of the exact meaning of the word "quantize" in Barclay, it is not reverting or converting the cepstral features back speech. Therefore, Barclay's quantized cepstral features are not speech, and these features are definitely not speech in raw uncompressed audio format. It is these cepstral features, not speech, that may be buffered before sending to a server. Column 5, lines 49–50.

Therefore, Barclay *does not show or suggest anything remotely related to storing of raw speech* as recited in the invention. Barclay does not store any part of the speech upon input. After Barclay's client processes the speech, by extracting and quantizing some cepstral features, the client stored the quantized features and not the speech itself before sending to a server.

Meisel also does not show any storing of raw speech for speech recognition or other analysis applications. Meisel describes storing of only enrollment data (column 4, line 67–column 5, line 9), but this enrollment data is not speech as recited in the invention. Rather, the enrollment data is used only for generating speaker specific parameters. Column 3, lines 47–49; column 4, line 67–column 5, line 1; column 12, lines 4–9. The generated speaker specific parameters are also not speech. These parameters "representing the speaker's pitch, the frequency spectrum of the speech as a function of time, and certain measurements of the speech signal in the time-domain." Column 2, line 66–column 3, line 3. These parameters are integral part of the Meisel invention for preprocessing. Meisel, abstract. Meisel describes only a method for generating parameters for preprocessing and does not show or suggest storing post-enrollment speech, or the actual speech for use in preprocessing.

Therefore, the references, individually or in combination, do not show or suggest storing the audio speech in one of more buffers in a raw uncompressed audio format. For at least this additional reason, claim 1 and its dependents should be allowable. Claims 9 and 11 recite similar limitations as in claim 1, and these claims and their dependents should be allowable for at least similar reasons.

Data Communication

In the office action, the examiner states "it is well known to buffer *data communications* both upon reception and before transmission to permit processing." However, claim 1 recites that a client *stores "the audio speech in one or more buffers* in a raw uncompressed audio format," which something very different from what the examiner states. Data communication refers to communication between two nonhuman devices, such as two computers. *Claim 1 does not refer to data communication, but storing of audio speech from for example, a human user.*

For at least this additional reason, the examiner has not made a *prima facie* case of obviousness, and claim 1 and its dependent should be allowable. Claims 9 and 11 recite similar limitations as in claim 1, and these claims and their dependents should be allowable for at least similar reasons.

No Encoding of the Received Speech

Additionally, claim 1 recites a client having a capability to "*encode* a buffer of the received audio speech before all of the audio speech is received." None of the cited references, individually or in combination, show or suggest this limitation.

Barclay does not teach or suggest encoding audio speech for transmission through a communication network, but rather describes extracting and quantizing cepstral features. Only the cepstral features are sent to a server through a communication network. Column 4, lines 3–5. As discussed above, cepstral features are *not speech*. And furthermore, the extracting or quantizing of Barclay are *not "encoding."*

Extracting features from speech is not encoding the speech. Barclay describes extracting features from speech by applying some mathematical operations on the speech to form features. Column 2, lines 3–8. These features are mathematical formations from speech without the speech content and quality. Column 2, lines 16–23. Extracting is unlike encoding because extracting does not preserve the content and quality of the speech.

Quantizing features is also not encoding speech. Barclay itself distinguishes "quantizing" from "encoding." Barclay's only use of the encoding term is at column 5, lines 61–64. There, Barclay describes encoding an end-of-speech (EOS) signal before sending it to the server. This does not show or suggest the recited invention because the *EOS signal is not speech*. Rather, the EOS signal is generated by the Barclay client after "a period of silence is encountered." Column

7, lines 23–24. As shown in figure 2A, the Barclay client waits for the user to stop speaking in step 36, and then the client (not a person) generates the EOS signal in step 38.

In all other instances in the reference, Barclay uses the term quantize (i.e., 26 times), and it is cepstral features that are quantized. Barclay never describes the cepstral features as being encoded. Moreover, as has been discussed, the cepstral features are not speech.

Barclay clearly *does not teach or suggest encoding* a buffer of the received audio speech before all of the audio speech is received. Meisel also *does not teach or suggest encoding* a buffer of the received audio speech. The references do not provide the features of benefits of the present invention.

The present invention encodes a buffer of the audio speech so that the server can evaluate the audio speech. For example, an embodiment of the invention evaluates the pronunciation accuracy of the audio speech. Such a system may be used to help speakers with a nonnative accent to learn to speak without the accent. The prior art does not show or suggest a system of the invention. For at least this additional reason, claim 1 and its dependents should be allowable. Claims 9 and 11 recite similar limitations as in claim 1, and these claims and their dependents should be allowable for at least similar reasons.

Conclusion

For the above reasons, applicant believes all claims now pending in this application are in condition for allowance. Applicant respectfully requests that a timely Notice of Allowance be issued in this case. If the examiner believes a telephone conference would expedite prosecution of this application, please contact the signee.

Respectfully submitted,

Aka Chan LLP

/Melvin D. Chan/

Melvin D. Chan
Reg. No. 39,626

Aka Chan LLP
900 Lafayette Street, Suite 710
Santa Clara, CA 95050
Tel: (408) 701-0035
Fax: (408) 608-1599
E-mail: mel@akachanlaw.com